Project 1

Ruiyu ZHANG

9/20/2022

Topic

In this project, the topic is Exploratory Data Analysis of History of Philosophy

Project title: Relationships of Philosophy Sentences and Other Factors. This Project is conducted by Ruiyu Zhang

Project Summary: This report analyzed The History of Philosophy dataset for the purpose of finding potential relationships between philosophy sentence lengths and factors like publication date, author, and school. Along the way, interesting findings on the history of philosophy are also reported. And, at the end stage, relationships of philosophy sentences with different authors and their frequencies are represented by wordclouds plots to produce a clearer view of frequency of words in sentences.

Method

I will follow the following steps: 1. Import the data 2. Clean the data 3. Process the data 4. Visualize the data 5. Conclude the findings

1. Import the data

First of all, I imported the dataset downloaded from Kaggle. By importing and explore the first look of the data, I see that there are 11 total columns, with only three columns of double types and the rest are character types.

```
data <- read.csv("~/Downloads/COLUMBIA UNIVERSITY/GR5243/philosophy_data.csv")
#head(data,6)
str(data)
```

```
## 'data.frame':    360808 obs. of  11 variables:
##  $ title                    : chr  "Plato - Complete Works" "Plato - Complete Works" "Plato - Comple
##  $ author                   : chr  "Plato" "Plato" "Plato" "Plato" ...
##  $ school                   : chr  "plato" "plato" "plato" "plato" ...
##  $ sentence_spacy           : chr  " What's new, Socrates, to make you leave your usual haunts in th
##  $ sentence_str             : chr  " What's new, Socrates, to make you leave your usual haunts in th
##  $ original_publication_date: int  -350 -350 -350 -350 -350 -350 -350 -350 -350 -350 ...
##  $ corpus_edition_date      : int  1997 1997 1997 1997 1997 1997 1997 1997 1997 1997 ...
##  $ sentence_length          : int  125 69 74 21 101 34 43 35 33 147 ...
##  $ sentence_lowered         : chr  " what's new, socrates, to make you leave your usual haunts in th
##  $ tokenized_txt            : chr  "['what', 'new', 'socrates', 'to', 'make', 'you', 'leave', 'your'
##  $ lemmatized_str           : chr  "   what be new , Socrates , to make -PRON- leave -PRON- usual hau
```

```
library("tidyverse")
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
```

```
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## -- Conflicts --------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(devtools)
```

```
## Loading required package: usethis
```

```
library(ggplot2)
library(dplyr)
library(tidyr)
library(ggpubr)
library(wordcloud)
```

```
## Loading required package: RColorBrewer
```

```
library(RColorBrewer)
library(wordcloud2)
library(tm)
```

```
## Loading required package: NLP
##
## Attaching package: 'NLP'
##
## The following object is masked from 'package:ggplot2':
##
##     annotate
```

2. Clean the data

Data is already pre-processed before downloading from Kaggle. The only data cleaning I need to do is to subset the data. Here, since my interest in this report is to explore relationships between philosophy sentence lengths and factors like publication date, author, and school. I subset the data to 6 columns of what I needed.

```
data <- data[, -c(4:5,9:11)]
head(data,6)
```

```
##                    title author school original_publication_date
## 1 Plato - Complete Works  Plato  plato                       -350
## 2 Plato - Complete Works  Plato  plato                       -350
## 3 Plato - Complete Works  Plato  plato                       -350
## 4 Plato - Complete Works  Plato  plato                       -350
## 5 Plato - Complete Works  Plato  plato                       -350
## 6 Plato - Complete Works  Plato  plato                       -350
##   corpus_edition_date sentence_length
## 1                1997             125
## 2                1997              69
## 3                1997              74
## 4                1997              21
## 5                1997             101
## 6                1997              34
```

```
colnames(data)
```

```
## [1] "title"                    "author"
## [3] "school"                   "original_publication_date"
## [5] "corpus_edition_date"      "sentence_length"
```

3. Process and mine the data

In this section, I will first evaluate the means of sentence length grouped by schools and authors.

```
data %>%
  group_by(school,author) %>%
  summarise_at(vars(sentence_length),
               list(sentence_length_mean=mean))
```

```
## # A tibble: 36 x 3
## # Groups:   school [13]
##    school     author       sentence_length_mean
##    <chr>      <chr>                       <dbl>
##  1 analytic   Kripke                      119.
##  2 analytic   Lewis                       110.
##  3 analytic   Moore                       167.
##  4 analytic   Popper                      140.
##  5 analytic   Quine                       122.
##  6 analytic   Russell                     146.
##  7 analytic   Wittgenstein                 84.9
##  8 aristotle  Aristotle                   153.
##  9 capitalism Keynes                      197.
## 10 capitalism Ricardo                     186.
## # ... with 26 more rows
```

From the above findings, I see that means of sentence lengths from each group of (school, author) are fairly around 100 to 200; in other words, there is no big difference among groups. It is worth to notice that (analytic, Wittgenstein) has the lowest mean of sentence length, which is only 84.884.

Next, let's evaluate data grouped by school, author, and title.

```
data %>%
  group_by(school,author,title) %>%
  summarise_at(vars(sentence_length),
               list(sentence_length_mean=mean))
```

```
## # A tibble: 59 x 4
## # Groups:   school, author [36]
##    school   author       title                              sentence_length_mean
##    <chr>    <chr>        <chr>                                             <dbl>
##  1 analytic Kripke       Naming And Necessity                              121.
##  2 analytic Kripke       Philosophical Troubles                            119.
##  3 analytic Lewis        Lewis - Papers                                    110.
##  4 analytic Moore        Philosophical Studies                             167.
##  5 analytic Popper       The Logic Of Scientific Discovery                 140.
##  6 analytic Quine        Quintessence                                      122.
##  7 analytic Russell      The Analysis Of Mind                              143.
##  8 analytic Russell      The Problems Of Philosophy                        155.
##  9 analytic Wittgenstein On Certainty                                       79.4
## 10 analytic Wittgenstein Philosophical Investigations                       83.6
## # ... with 49 more rows
```

Now, when I dive in deeper of the dataset by grouping by school, author, and title, I see three groups with

lower than 100 means: (analytic,Wittgenstein,On Certainty),(analytic,Wittgenstein,Philosophical Investigations),(nietzsche,Nietzsche, Thus Spake Zarathustra).

Since one of the goals is to examine the relationships between sentence length with other factors. Let's check the correlation coefficients:

```
cor(data$sentence_length,data$original_publication_date)
```

```
## [1] 0.06739653
```

```
cor(data$sentence_length,data$corpus_edition_date)
```

```
## [1] 0.01663432
```

By above correlation coefficient, we see that both sentence length and corpus edition date and sentence length and original publication date is actually weakly correlated. In other words, we can not find a strong positive or negative relationship between those two factors. After examining the relationship between two numerical features. Let's process Kruskal-Wallis test to find if there is any relationship between categorical features, such as school, author, title with sentence length. The null hypothesis will be the medians of groups are equal, and the alternative hypothesis is that at least one of the group has different median.

```
kruskal.test(sentence_length ~ school, data = data)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  sentence_length by school
## Kruskal-Wallis chi-squared = 22436, df = 12, p-value < 2.2e-16
```

```
kruskal.test(sentence_length ~ author, data = data)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  sentence_length by author
## Kruskal-Wallis chi-squared = 31272, df = 35, p-value < 2.2e-16
```

```
kruskal.test(sentence_length ~ title, data = data)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  sentence_length by title
## Kruskal-Wallis chi-squared = 34909, df = 58, p-value < 2.2e-16
```

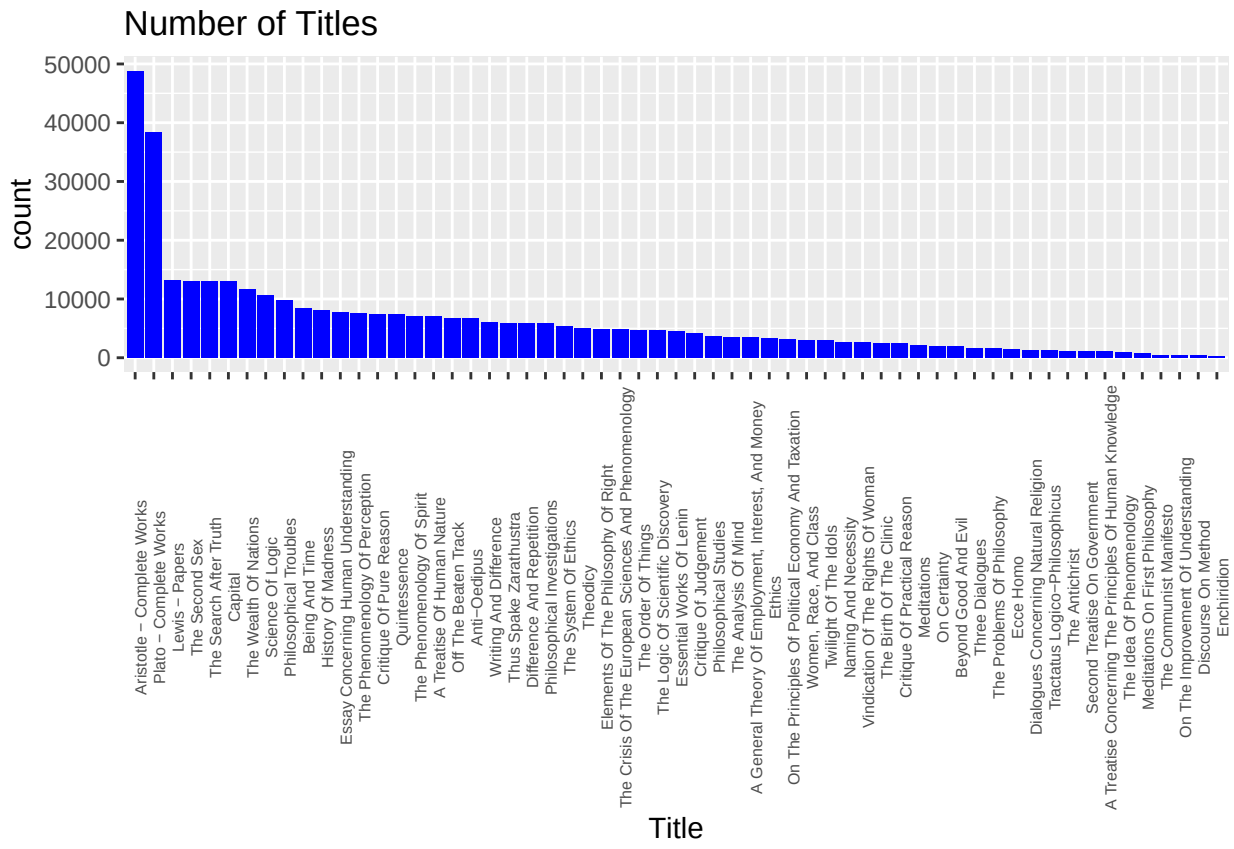
As the p-values for all three kw rank sum test are less than the significance level 0.05, we reject the null and conclude that medians of sentence lengths are diffrent from group schoo, author, and title. By now, we have examine both numerical and categorical relationship with sentence length. We find that there is no particular correlation between original publication date and sentence length and edition date to sentence length. But, it is worth to mention that from KW test we found medians of sentence lengths are different among schools, authors, titles.

4. Visualize the data

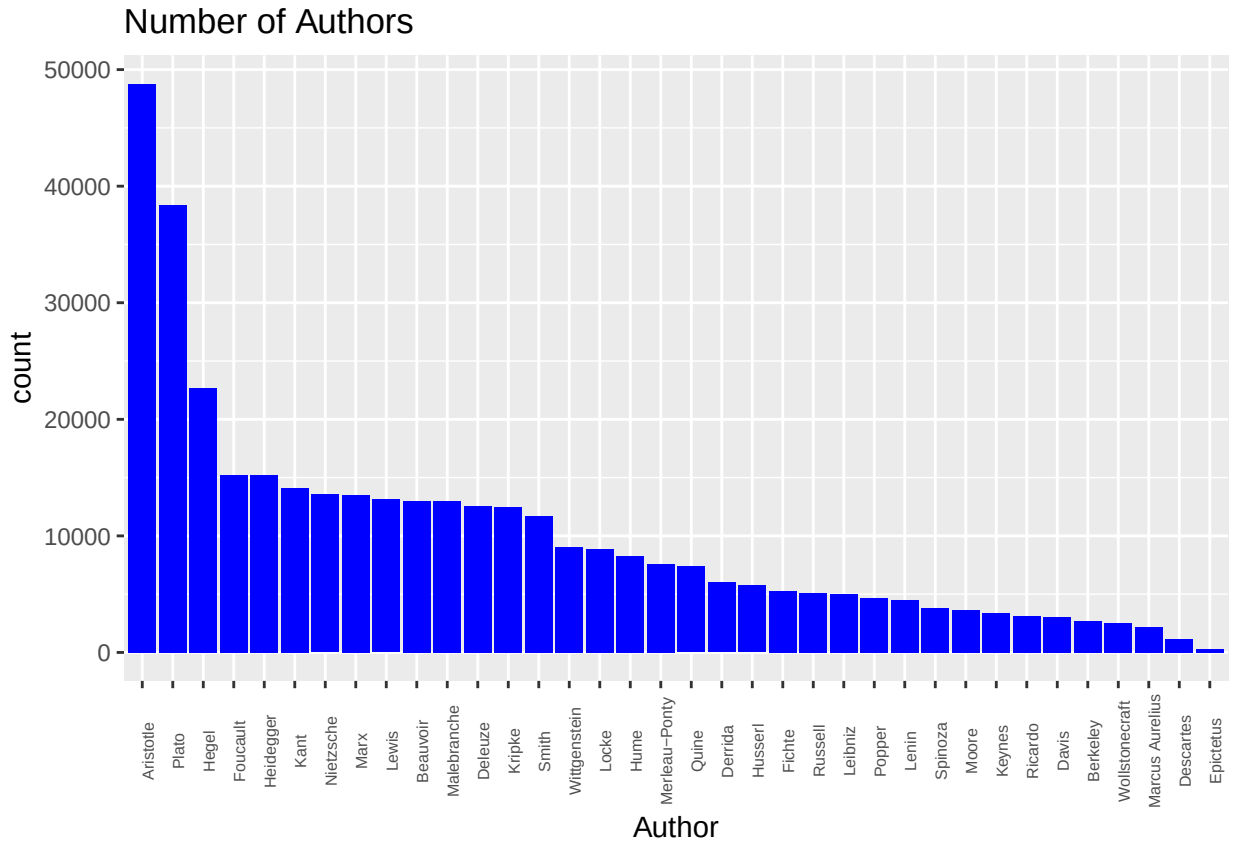
Let's take a closer look of the dataset with Categorical Features:

```
ggplot(data= data, aes(x= fct_infreq(title))) +
  geom_bar(fill="blue")+
  labs(x="Title")+
```

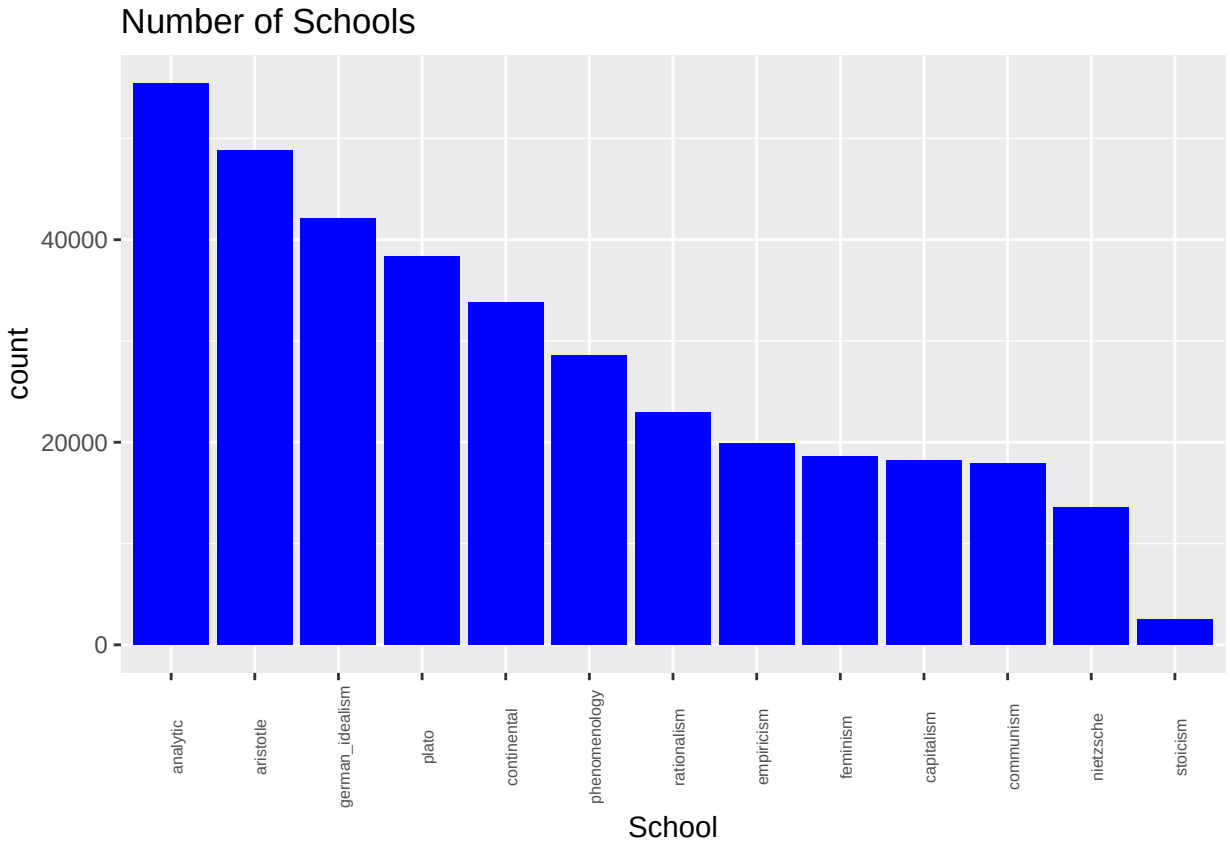
```
  ggtitle("Number of Titles")+
  theme(axis.text.x=element_text(size=6,angle=90))
```




```
#author
ggplot(data= data, aes(x= fct_infreq(author))) +
  geom_bar(fill="blue")+
  labs(x="Author")+
  ggtitle("Number of Authors")+
  theme(axis.text.x=element_text(size=6,angle=90))
```




```
#school
ggplot(data= data, aes(x= fct_infreq(school))) +
  geom_bar(fill="blue")+
  labs(x="School")+
  ggtitle("Number of Schools")+
  theme(axis.text.x=element_text(size=6,angle=90))
```

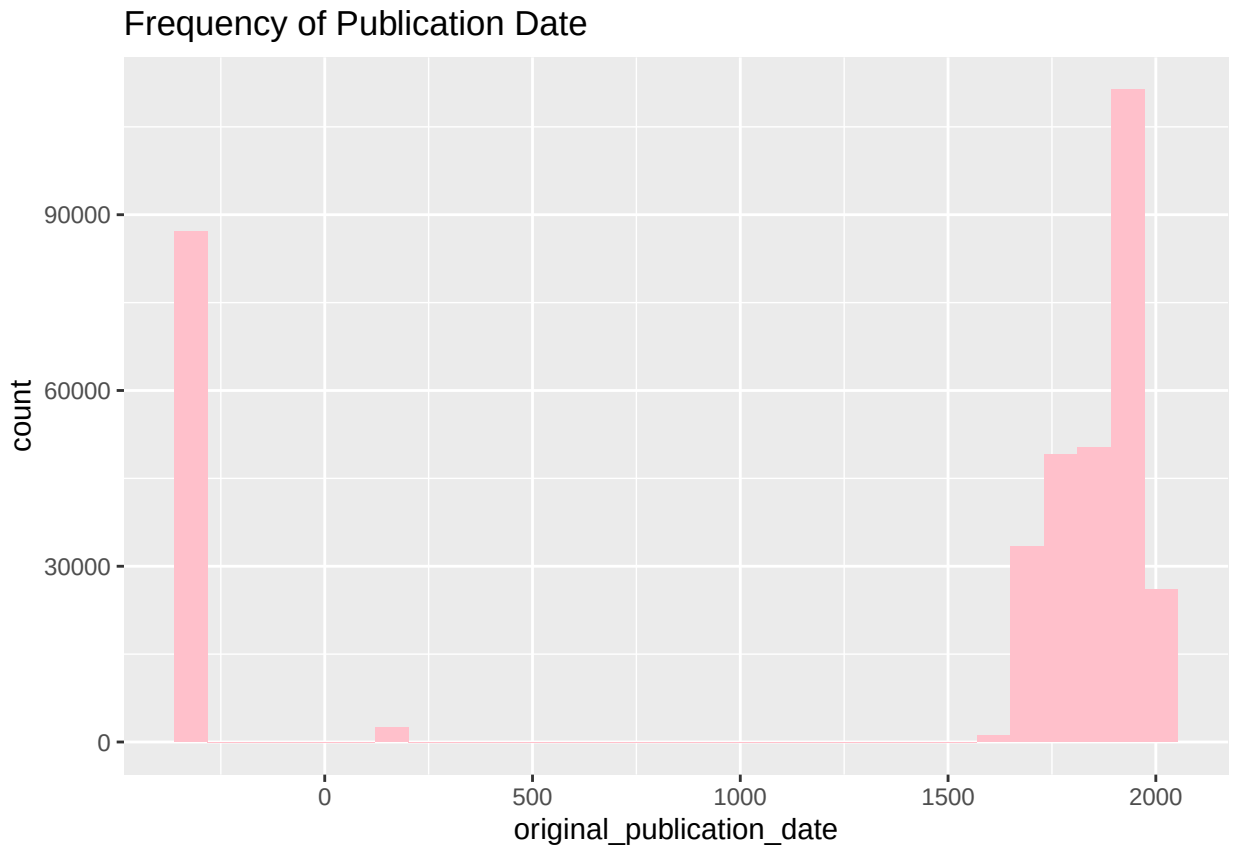


From the above categorical data plots, I am able to visualize the counts of different title, author, and school in the data. As for title, we have the most entries from Aristotle-Complete Works, following by Plato-Complete Works on the second. As for author, we have the similar result, where Aristotle ranked the first and Plato the second. As for number of school, we have Analytic the most and Aristotle the second.

Next, let's take a closer look at the dataset with numerical features:

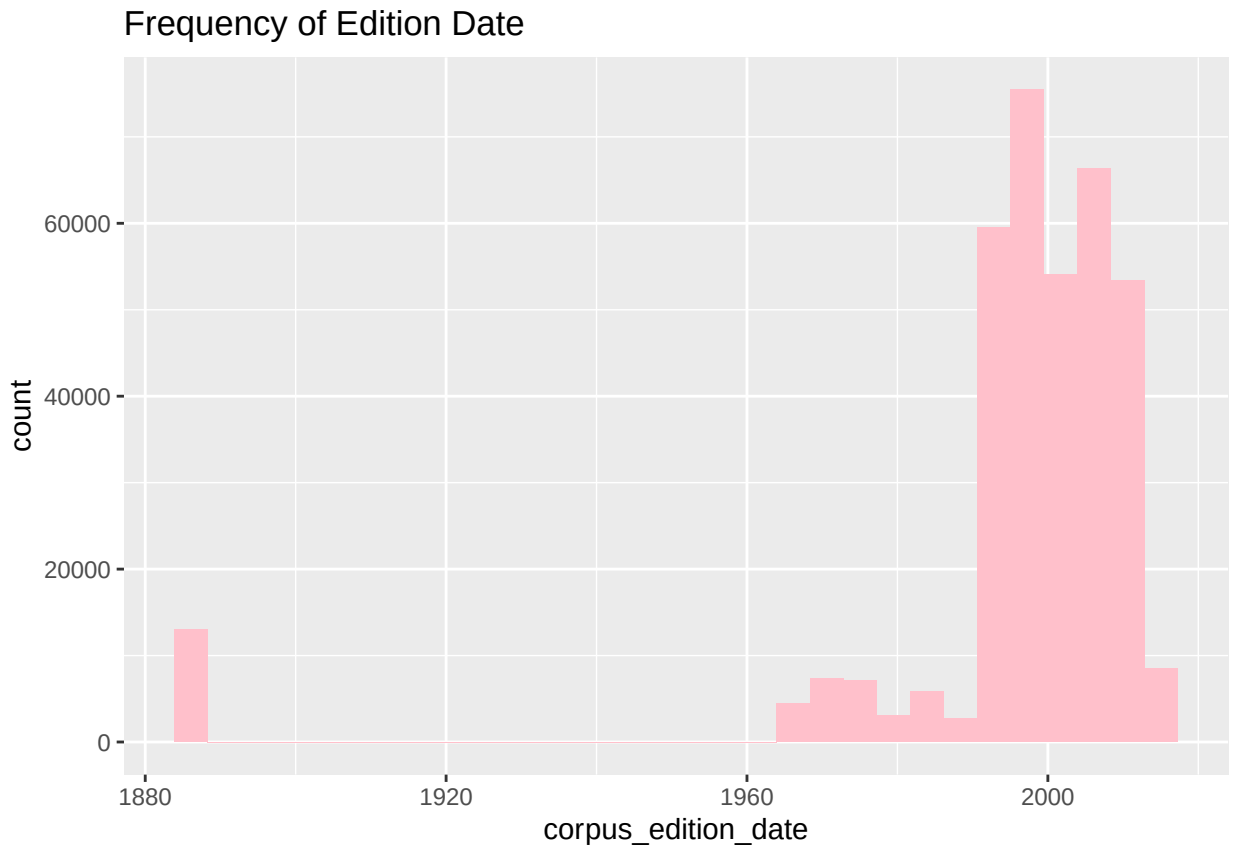
```
#Original Publication Date
ggplot(data = data, aes(x=original_publication_date)) +
  geom_histogram(fill="pink")+
  ggtitle("Frequency of Publication Date")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
#edition Date
ggplot(data = data, aes(x=corpus_edition_date)) +
  geom_histogram(fill="pink")+
  ggtitle("Frequency of Edition Date")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

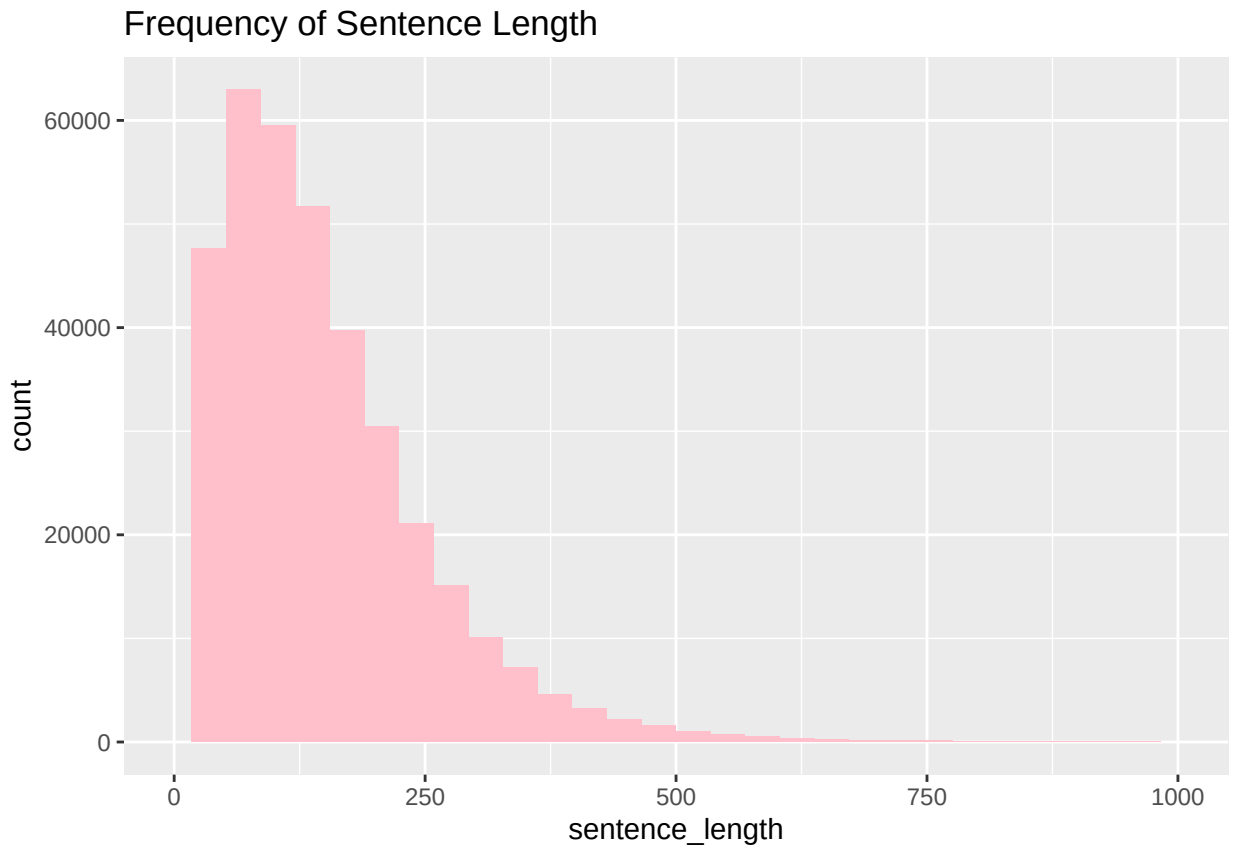



```
#Sentence Length
ggplot(data = data, aes(x=sentence_length)) +
  geom_histogram(fill="pink")+
  xlim(0,1000)+
  ggtitle("Frequency of Sentence Length")
```

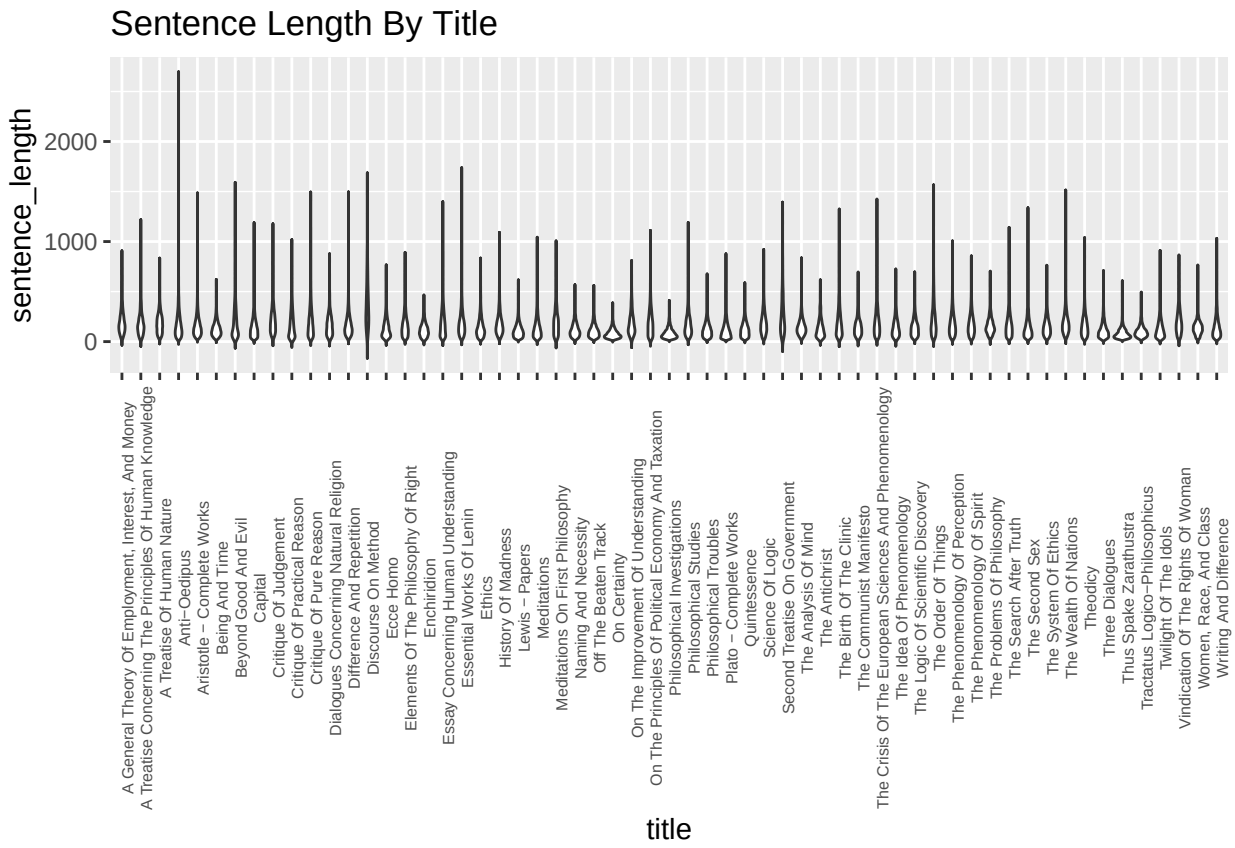
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 83 rows containing non-finite values (stat_bin).
```

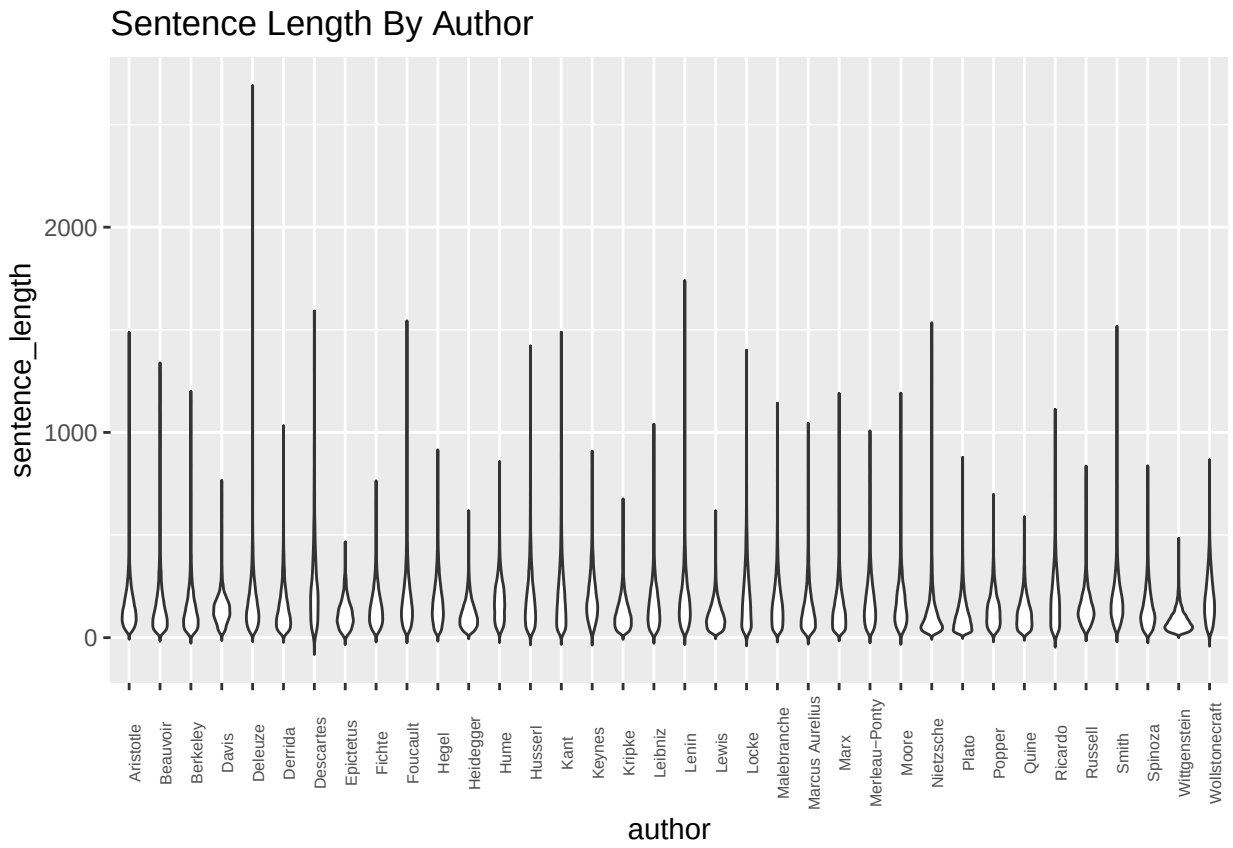
```
## Warning: Removed 2 rows containing missing values (geom_bar).
```



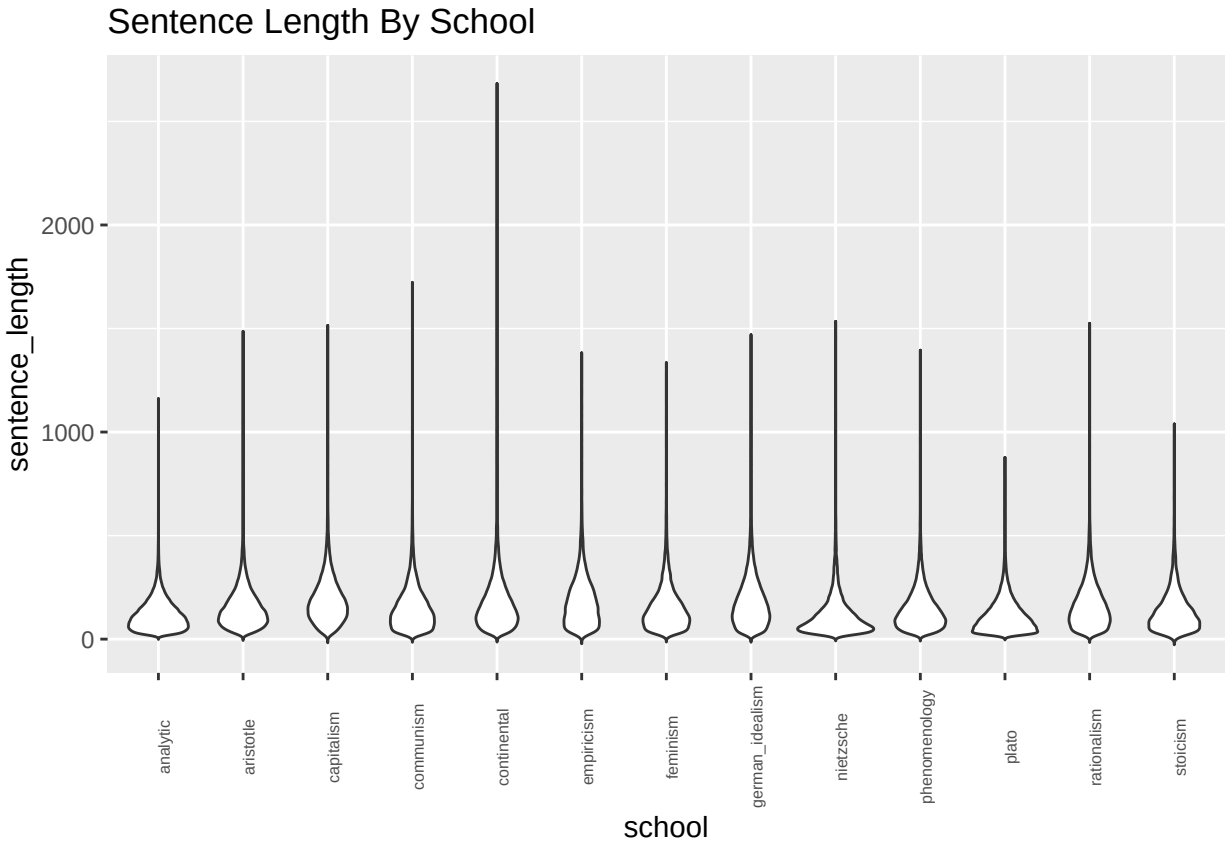

```
#Sentence Length By School
#ggplot(data = data, aes(x=school,y=sentence_length)) +
#  geom_histogram(aes(fill=school))+
#  xlim(0,1000)+
#  ggtitle("Frequency of Sentence Length By School")

#Sentence Length By Title
ggplot(data = data, aes(x=title, y=sentence_length)) +
  geom_violin(trim=F)+
  ggtitle("Sentence Length By Title")+
  theme(axis.text.x=element_text(size=6,angle=90))
```




```
#Sentence Length By Author
ggplot(data = data, aes(x=author, y=sentence_length)) +
  geom_violin(trim=F)+
  ggtitle("Sentence Length By Author")+
  theme(axis.text.x=element_text(size=6,angle=90))
```




```
#Sentence Length By School
ggplot(data = data, aes(x=school, y=sentence_length)) +
  geom_violin(trim=F)+
  ggtitle("Sentence Length By School")+
  theme(axis.text.x=element_text(size=6,angle=90))
```



From the above numerical data plots, I am able to visualize most of philosophy sentences were originally published from 1500 to 2000, edited from 1980 to 2020. I see a right-skewed histogram of sentences length, most of sentences length go from 0 - 250.

5. Word Clouds

In this section, I will examine the most frequent words by word clouds and frequency barplot. Since the whole dataset is too large and unable to generate word clouds due to vector memory reasons. I will subset the data by the author from "Number of Authors" plot developed in part 4. They will be Plato, Hegel, and Foucault.

I skipped the Aristotle with most counts because it is unable to run because of vector memory reason due to subset too large. The following is Plato subset:

```
#Plato
#reload the data and clean the data
library(tm)
data1 <- read.csv("~/Downloads/COLUMBIA UNIVERSITY/GR5243/philosophy_data.csv")
data1 <- data1 %>%
  filter(author == "Plato")

text <- data1$sentence_lowered
docs <- Corpus(VectorSource(text))
#inspect(docs)
toSpace <- content_transformer(function (x , pattern ) gsub(pattern, " ", x))
docs <- tm_map(docs, toSpace, "/")
```

```
## Warning in tm_map.SimpleCorpus(docs, toSpace, "/"): transformation drops
## documents
```

```
docs <- tm_map(docs, toSpace, "@")
```

```
## Warning in tm_map.SimpleCorpus(docs, toSpace, "@"): transformation drops
## documents
```

```
docs <- tm_map(docs, toSpace, "\\|")
```

```
## Warning in tm_map.SimpleCorpus(docs, toSpace, "\\|"): transformation drops
## documents
```

```
# Convert the text to lower case
docs <- tm_map(docs, content_transformer(tolower))
```

```
## Warning in tm_map.SimpleCorpus(docs, content_transformer(tolower)):
## transformation drops documents
```

```
# Remove numbers
docs <- tm_map(docs, removeNumbers)
```

```
## Warning in tm_map.SimpleCorpus(docs, removeNumbers): transformation drops
## documents
```

```
# Remove english common stopwords
docs <- tm_map(docs, removeWords, stopwords("english"))
```

```
## Warning in tm_map.SimpleCorpus(docs, removeWords, stopwords("english")):
## transformation drops documents
```

```
# Remove punctuations
docs <- tm_map(docs, removePunctuation)
```

```
## Warning in tm_map.SimpleCorpus(docs, removePunctuation): transformation drops
## documents
```

```
# Eliminate extra white spaces
docs <- tm_map(docs, stripWhitespace)
```

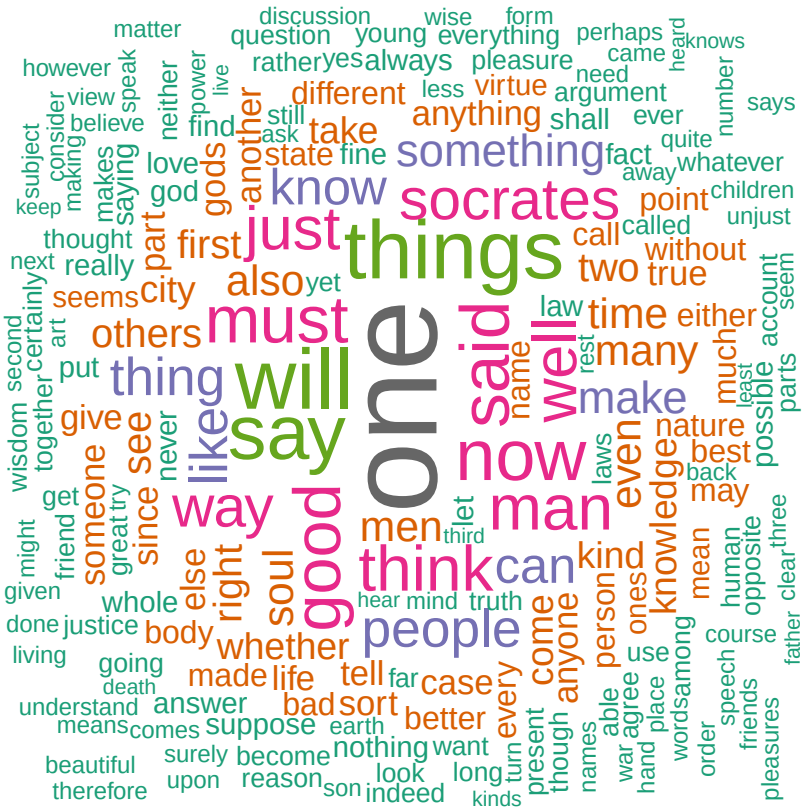
```
## Warning in tm_map.SimpleCorpus(docs, stripWhitespace): transformation drops
## documents
```

```
dtm <- TermDocumentMatrix(docs)
m <- as.matrix(dtm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
head(d, 10)
```

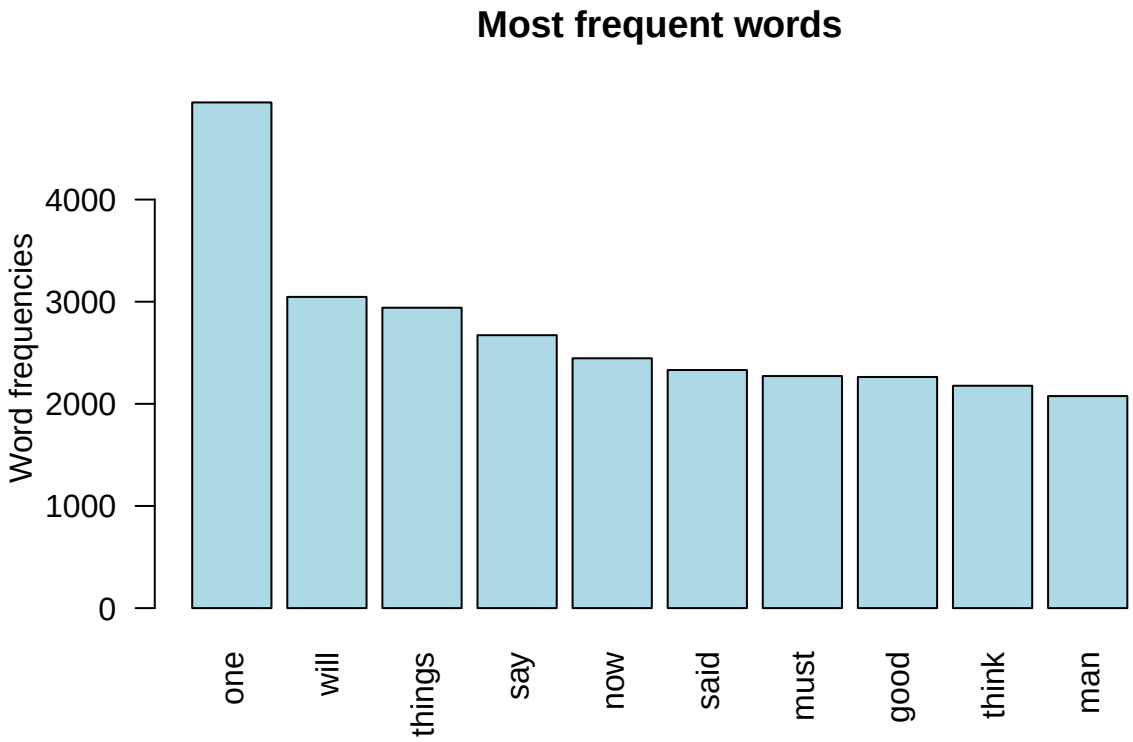
```
##          word freq
## one       one 4951
## will     will 3047
## things things 2941
## say       say 2672
## now       now 2446
## said     said 2331
## must     must 2272
## good     good 2263
## think   think 2177
## man       man 2076
```

```
set.seed(1234)
wordcloud(words = d$word, freq = d$freq, min.freq = 1,
```

```
          max.words=200, random.order=FALSE, rot.per=0.35,
          colors=brewer.pal(8, "Dark2"))
```




```
barplot(d[1:10,]$freq, las = 2, names.arg = d[1:10,]$word,
        col ="lightblue", main ="Most frequent words",
        ylab = "Word frequencies")
```



Most frequent words for Plato are “one”, “will”, and “things”.

The following is Hegel subset:

```
#reload the data and clean the data
data1 <- read.csv("~/Downloads/COLUMBIA UNIVERSITY/GR5243/philosophy_data.csv")
data1 <- data1 %>%
  filter(author== "Hegel")

text <- data1$sentence_lowered
docs <- Corpus(VectorSource(text))
#inspect(docs)
toSpace <- content_transformer(function (x , pattern ) gsub(pattern, " ", x))
docs <- tm_map(docs, toSpace, "/")
```

```
## Warning in tm_map.SimpleCorpus(docs, toSpace, "/"): transformation drops
## documents
```

```
docs <- tm_map(docs, toSpace, "@")
```

```
## Warning in tm_map.SimpleCorpus(docs, toSpace, "@"): transformation drops
## documents
```

```
docs <- tm_map(docs, toSpace, "\\|")
```

```
## Warning in tm_map.SimpleCorpus(docs, toSpace, "\\|"): transformation drops
## documents
```

```
# Convert the text to lower case
docs <- tm_map(docs, content_transformer(tolower))
```

```
## Warning in tm_map.SimpleCorpus(docs, content_transformer(tolower)):
## transformation drops documents
```

```r
# Remove numbers
docs <- tm_map(docs, removeNumbers)
```

```
## Warning in tm_map.SimpleCorpus(docs, removeNumbers): transformation drops
## documents
```

```r
# Remove english common stopwords
docs <- tm_map(docs, removeWords, stopwords("english"))
```

```
## Warning in tm_map.SimpleCorpus(docs, removeWords, stopwords("english")):
## transformation drops documents
```

```r
# Remove punctuations
docs <- tm_map(docs, removePunctuation)
```

```
## Warning in tm_map.SimpleCorpus(docs, removePunctuation): transformation drops
## documents
```

```r
# Eliminate extra white spaces
docs <- tm_map(docs, stripWhitespace)
```

```
## Warning in tm_map.SimpleCorpus(docs, stripWhitespace): transformation drops
## documents
```

```r
dtm <- TermDocumentMatrix(docs)
m <- as.matrix(dtm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
head(d, 10)
```

```
##                         word freq
## one                      one 3198
## self                    self 3183
## consciousness consciousness 2479
## concept             concept 1958
## therefore         therefore 1953
## also                   also 1799
## form                   form 1798
## existence         existence 1771
## content             content 1679
## first                 first 1630
```

```r
set.seed(1234)
wordcloud(words = d$word, freq = d$freq, min.freq = 1,
          max.words=200, random.order=FALSE, rot.per=0.35,
          colors=brewer.pal(8, "Dark2"))
```

```
## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : determinations could not be fit on page. It will not be plotted.
```

```
## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : negative could not be fit on page. It will not be plotted.
```

```
## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : simple could not be fit on page. It will not be plotted.
```

```
## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : general could not be fit on page. It will not be plotted.
```

```
## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
```

```
## 200, : thought could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : determined could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : particular could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : difference could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : nothing could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : moment could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : matter could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : abstract could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : reality could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : ground could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : merely could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : universality could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : present could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : concrete could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : essential could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : substance could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : action could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : actuality could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : notion could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : whole could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : equally could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : negation could not be fit on page. It will not be plotted.
```

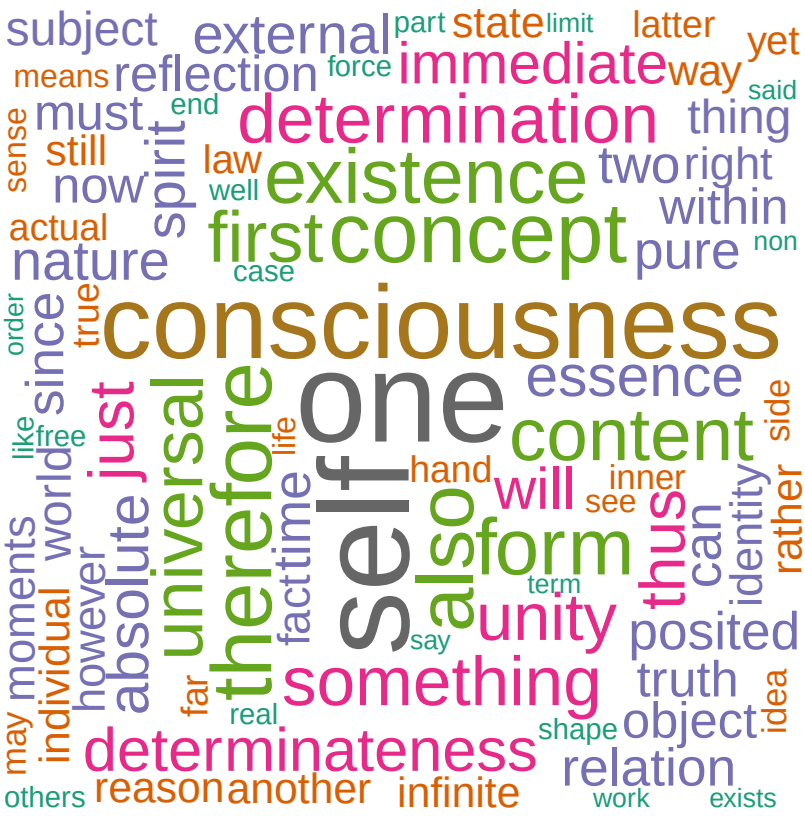
```
## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : movement could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : quantum could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : immediately could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : even could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : connection could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : power could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : determinate could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : taken could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : already could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : immediacy could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : reference could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : indifferent could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : judgment could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : principle could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : subjective could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : finite could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : without could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : objective could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : positive could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : point could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : individuality could not be fit on page. It will not be plotted.
```

```
## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : knowledge could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : things could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : purpose could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : made could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : contains could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : ethical could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : necessity could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : science could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : becomes could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : given could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : good could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : consequently could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : freedom could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : element could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : essentially could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : hence could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : contrary could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : opposite could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : cognition could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : specific could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : become could not be fit on page. It will not be plotted.
```

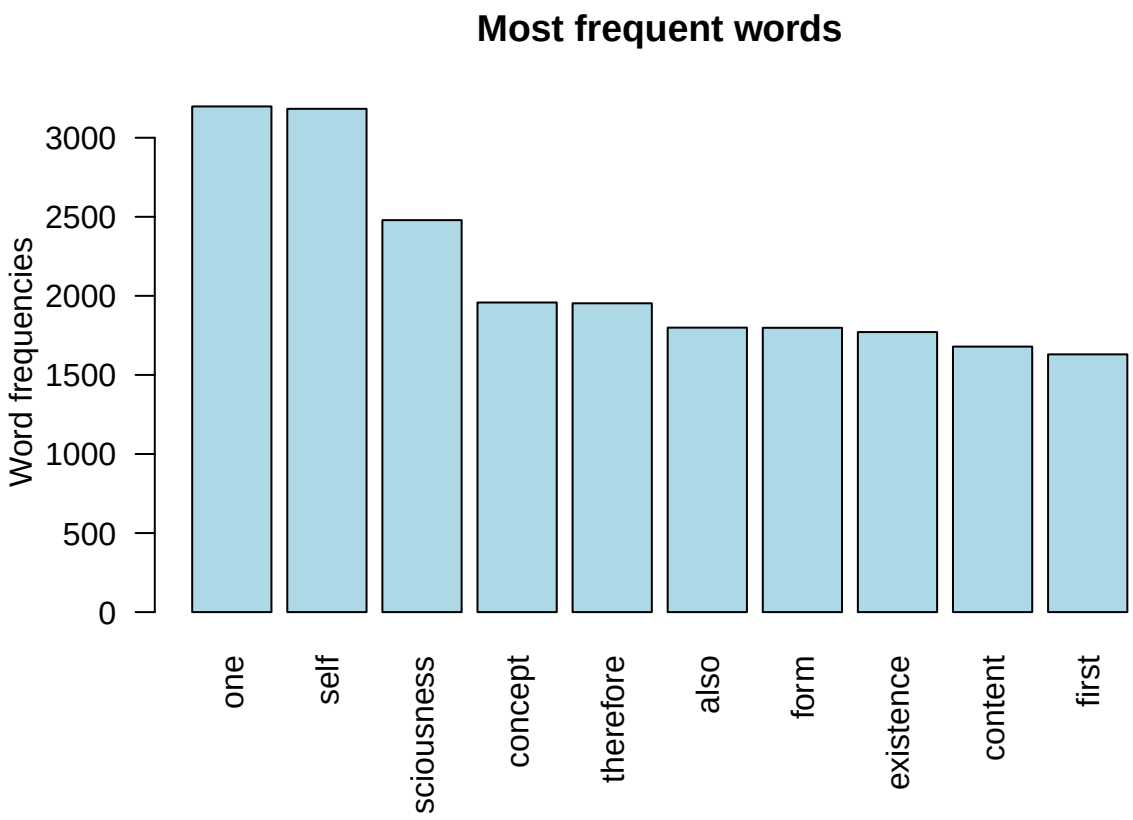
```
## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : thereby could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : result could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : second could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : positedness could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : constitutes could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : different could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : syllogism could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : back could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : property could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : mediation could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : externality could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : longer could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : sublated could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : called could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : activity could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : opposition could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : identical could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : appearance could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : remains could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : totality could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : indeed could not be fit on page. It will not be plotted.
```

```
## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : subsistence could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : understanding could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : qualitative could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : aspect could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : much could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : formal could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : mere could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : whether could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : outside could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : reflected could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : makes could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : simply could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : process could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : negativity could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : supposed could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : namely could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : contradiction could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : sphere could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : precisely could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : beyond could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : objectivity could not be fit on page. It will not be plotted.
```

```
## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : appears could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : according could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : whose could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : magnitude could not be fit on page. It will not be plotted.
```




```
barplot(d[1:10,]$freq, las = 2, names.arg = d[1:10,]$word,
        col ="lightblue", main ="Most frequent words",
        ylab = "Word frequencies")
```



Most frequent words for Plato are “one”, “self”, “sciousness”.

The following is Foucault subset:

```
#Plato
#reload the data and clean the data
library(tm)
data1 <- read.csv("~/Downloads/COLUMBIA UNIVERSITY/GR5243/philosophy_data.csv")
data1 <- data1 %>%
  filter(author == "Foucault")

text <- data1$sentence_lowered
docs <- Corpus(VectorSource(text))
#inspect(docs)
toSpace <- content_transformer(function (x , pattern ) gsub(pattern, " ", x))
docs <- tm_map(docs, toSpace, "/")
```

```
## Warning in tm_map.SimpleCorpus(docs, toSpace, "/"): transformation drops
## documents
```

```
docs <- tm_map(docs, toSpace, "@")
```

```
## Warning in tm_map.SimpleCorpus(docs, toSpace, "@"): transformation drops
## documents
```

```
docs <- tm_map(docs, toSpace, "\\|")
```

```
## Warning in tm_map.SimpleCorpus(docs, toSpace, "\\|"): transformation drops
## documents
```

```r
# Convert the text to lower case
docs <- tm_map(docs, content_transformer(tolower))
```

```
## Warning in tm_map.SimpleCorpus(docs, content_transformer(tolower)):
## transformation drops documents
```

```r
# Remove numbers
docs <- tm_map(docs, removeNumbers)
```

```
## Warning in tm_map.SimpleCorpus(docs, removeNumbers): transformation drops
## documents
```

```r
# Remove english common stopwords
docs <- tm_map(docs, removeWords, stopwords("english"))
```

```
## Warning in tm_map.SimpleCorpus(docs, removeWords, stopwords("english")):
## transformation drops documents
```

```r
# Remove punctuations
docs <- tm_map(docs, removePunctuation)
```

```
## Warning in tm_map.SimpleCorpus(docs, removePunctuation): transformation drops
## documents
```

```r
# Eliminate extra white spaces
docs <- tm_map(docs, stripWhitespace)
```

```
## Warning in tm_map.SimpleCorpus(docs, stripWhitespace): transformation drops
## documents
```

```r
dtm <- TermDocumentMatrix(docs)
m <- as.matrix(dtm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
head(d, 10)
```

```
##                word freq
## madness     madness 2043
## one             one 1973
## language   language 1136
## form           form 1012
## can             can 1011
## century     century  967
## nature       nature  916
## knowledge knowledge  897
## man             man  892
## order         order  803
```

```r
set.seed(1234)
wordcloud(words = d$word, freq = d$freq, min.freq = 1,
          max.words=200, random.order=FALSE, rot.per=0.35,
          colors=brewer.pal(8, "Dark2"))
```

```
## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : discourse could not be fit on page. It will not be plotted.
```

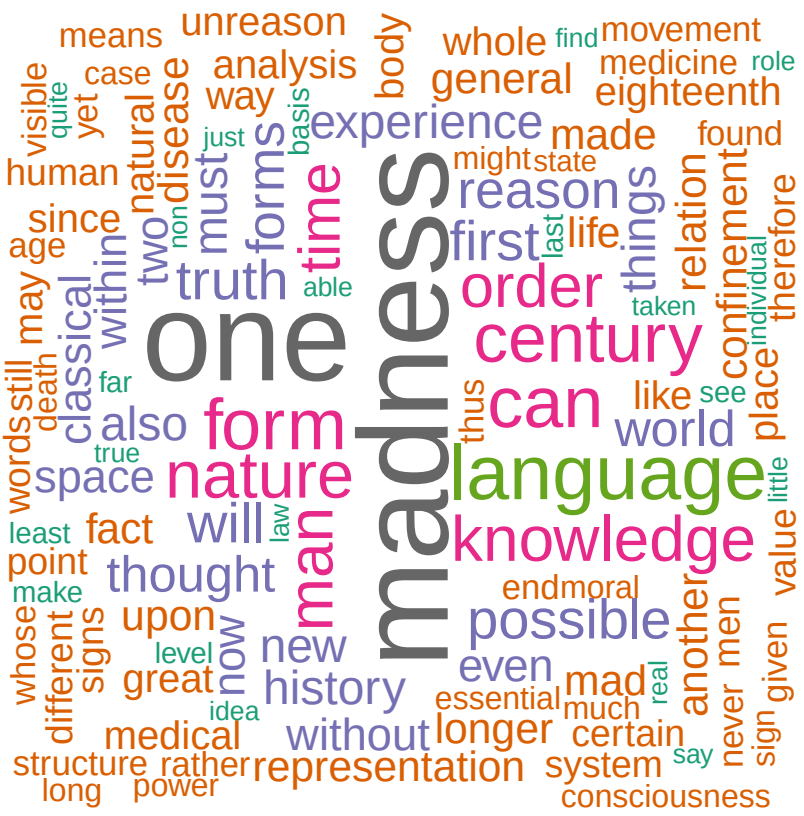
```
## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : become could not be fit on page. It will not be plotted.
```

```
## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : work could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : always could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : nothing could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : gaze could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : less could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : towards could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : became could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : together could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : object could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : sciences could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : subject could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : already could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : though could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : moment could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : question could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : elements could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : existence could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : hand could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : meaning could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : sense could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : part could not be fit on page. It will not be plotted.
```

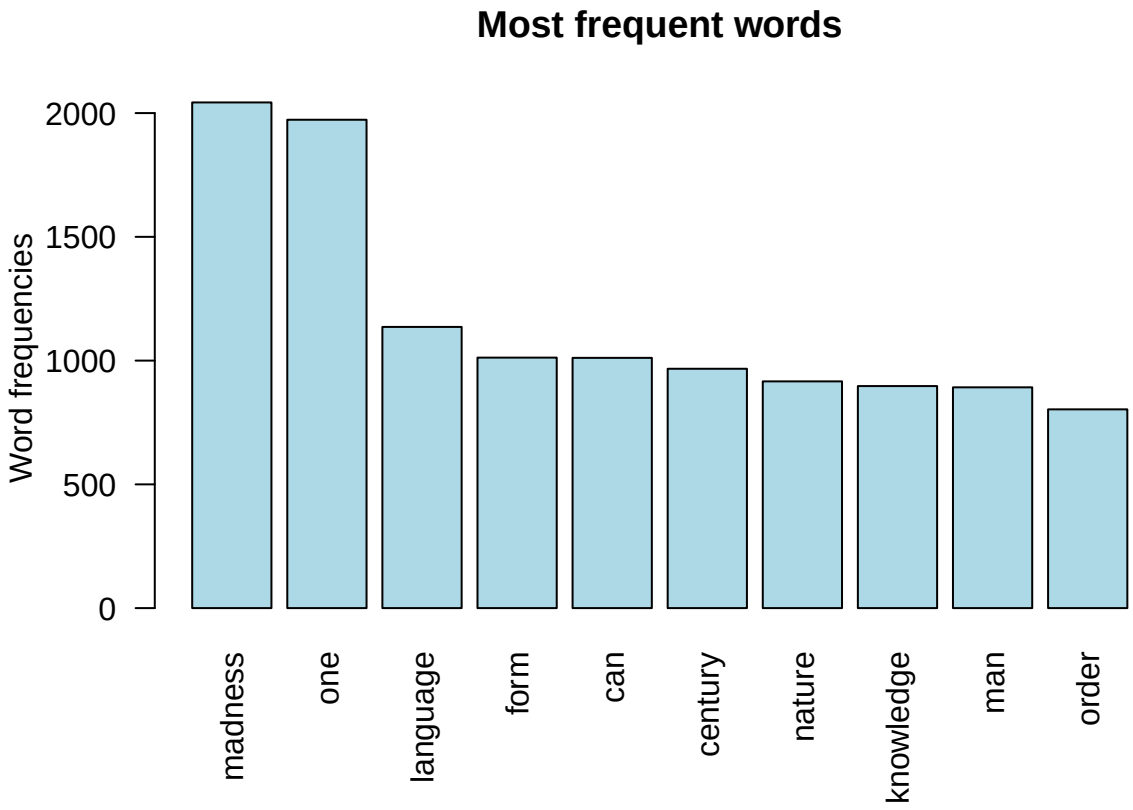
```
## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : paris could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : well could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : becomes could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : perhaps could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : possibility could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : nineteenth could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : necessary could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : others could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : wealth could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : social could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : many could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : common could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : take could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : particular could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : immediate could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : back could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : according could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : fundamental could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : madman could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : pathological could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : beings could not be fit on page. It will not be plotted.
```

```
## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : manner could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : insane could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : soul could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : use could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : makes could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : function could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : simple could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : people could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : origin could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : living could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : years could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : doubt could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : labour could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : hospital could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : science could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : simply could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : perception could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : mind could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : second could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : process could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : delirium could not be fit on page. It will not be plotted.
```

```
## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : series could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : society could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : day could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : character could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : species could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : know could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : element could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : diseases could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : result could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : linked could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : brought could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : seventeenth could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : pure could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : number could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : entirely could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : ever could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : theory could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : organic could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : word could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : light could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 1, max.words =
## 200, : positive could not be fit on page. It will not be plotted.
```




```
barplot(d[1:10,]$freq, las = 2, names.arg = d[1:10,]$word,
        col ="lightblue", main ="Most frequent words",
        ylab = "Word frequencies")
```



Most frequent words for Foucault are “madness”, “one”, “language”.

6. Conclusion

After all, there is no strong relationship found between sentence lengths of philosophy and original publication date, sentence length of philosophy and edition date. Though sentence length of philosophy are different with different means among authors, schools, and titles. However, we found interesting facts that means of sentence lengths are fairly around 100-200; also, as shown by violin plots titled by “sentence length by author”, “sentence length by title”, and “sentence length by school”, we can also visualize that sentence lengths are mostly in the range of 100 to 200. By the end, wordclouds were produced. I am able to see that different authors have different frequencies on using words: most frequent words for Plato are “one”, “self”, “sciousness”; most frequent words for Plato are “one”, “will”, and “things”; most frequent words for Foucault are “madness”, “one”, “language”.